International Journal of Research Publication and Reviews, Vol 5, no 6, pp 4578-4582 June 2024

WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Artificial Neural Networks in Speech Recogonization

Mohammed idris S

Student, Masters of Computer Applications, School of CS & IT, Jain (Deemed-To-Be-University), Bangalore, India,
mohammedidrissss3@gmail.com
DOI: https://doi.org/10.55248/gengpi.5.0624.1568

A B S T R A C T

Speech is the most effective form of interpersonal communication, which makes it an excellent choice for machine interfaces. This potential is shown by the growing need for automatic speech recognition systems. In this field, several approaches have been investigated, such as the Hidden Markov Model (HMM), Dynamic Time Warping (DTW), and Vector Quantization (VQ). However, the extensive use of artificial neural networks (ANN) in voice recognition is explored in this research. It carefully looks at the variety of neural network-based methods that can be used for speech recognition and compares their benefits and drawbacks. In the end, the study offers a recommendation for the best approach to further this field.

Keywords-–Neural Networks, Training Algorithm, Speech ,Recognition, Artificial Intelligence, Feature Extraction, Pattern ,Recognition, LPC, MFCC, Perceptron, Feedforward Neural Networks, etc.

INTRODUCTION:

Speech is still one of the easiest and most effective ways for people to communicate with each other. People learn to speak and communicate from an early age and rely on these abilities all their lives without receiving formal education. Speech is naturally preferred, even when interacting with machines, as users want a smooth and simple way to communicate. The interaction paradigm provided by speech is more intuitive and natural than that of heavy interfaces like keyboards and mouse.

But speaking is a difficult task because of the complicated vocal tract and articulatory mechanisms in humans, which are not under conscious control. The diversity and complexity of speech signals are influenced by a number of variables, including echoes, background noise, accents, articulation, pronunciation, emotional state, gender, pitch, speed, and loudness.. ASR, or recognition, is essential. Sophisticated algorithms and technologies are used by ASR to translate voice signals into comprehensible word sequences. The ideal ASR system should be able to identify speech straight from digital waveforms. As of right now, ASR systems are able to comprehend hundreds of phrases in practical settings.

This changing environment emphasizes how crucial it is for speech recognition technology to keep improving in order to further improve human-computer interactions.

PROBLEM STATEMENT:

Even with the great progress made in Artificial Neural Networks (ANNs), reliable and precise speech detection is still a difficult task. Attaining high performance levels in speech recognition systems is hampered by the complexity and variety of speech signals, which include variances in accents, pronunciation, and ambient noise. Even though ANNs have demonstrated promise in overcoming these obstacles, a number of important problems still need to be resolved in order to increase the efficacy and efficiency of speech recognition technology.

LITERATURE SURVEY:

Examining previous studies, research papers, and publications on the subject is part of a literature study for a research paper on Artificial Neural Networks (ANNs) in speech recognition. For such a study, the following is a sample format for a literature survey:

Literature Review: Speech Recognition Using Artificial Neural Networks

1. Overview of Neural Networks and Speech Recognition a succinct description of the importance of speech recognition technology in numerous applications.

An overview of artificial neural networks (ANNs) and how they work in tasks involving speech recognition.

2. A Review of Speech Recognition Techniques Throughout History

An examination of classic voice recognition methods like Vector Quantization (VQ), Dynamic Time Warping (DTW), and Hidden Markov Models (HMMs).

Talk about the drawbacks and difficulties with old approaches, which resulted in the introduction of methods based on neural networks.

3. Architectures of Neural Networks for Speech Recognition summary of the various neural network architectures for voice recognition, such as: Convolutional neural networks (CNNs) are utilized to extract features from various time-frequency representations, such as spectrograms.

For simulating temporal dependencies in voice sequences, recurrent neural networks (RNNs) like Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) are used.

CNNs and RNNs combined in hybrid architectures for sequence modeling and feature extraction. a comparison of the benefits and drawbacks of various neural network topologies for speech recognition applications.

4. Methods for Feature Extraction

Examining feature extraction techniques that are frequently combined with neural networks to recognize speech, including:

MFCCs, or Mel Frequency Cepstral Coefficients

LPC, or linear predictive coding

Spectral characteristics

Assessment of the efficacy of various methods for feature extraction to extract pertinent information from voice sounds.

5. Techniques for Training and Optimization

Backpropagation and gradient descent-based optimization algorithms are two examples of training techniques and optimization tactics discussed for neural network-based voice recognition systems.

strategies for regularization to avoid overfitting.

Model selection and adjustment of hyperparameters.

An examination of current developments in optimization methods and training algorithms to boost neural network models' output for speech recognition applications.

RELATED WORK:

1. Synopsis of the Literature in Existence

Give a succinct overview of the speech recognition industry and the advancements made possible by neural networks.

Emphasize the importance of earlier research initiatives in enhancing neural network-based voice recognition systems' robustness, accuracy, and efficiency.

2. Evaluation of Architectures for Neural Networks

Provide an overview of earlier research that looked at different neural network designs for speech recognition tasks, such as:

Spectrophotomically-based feature extraction using convolutional neural networks (CNNs).

For sequence modeling, use recurrent neural networks (RNNs) like Gated Recurrent Unit (GRU) or Long Short-Term Memory (LSTM).

CNNs and RNNs combined in hybrid architectures for end-to-end voice recognition.

Examine each architecture's advantages and disadvantages in light of the empirical findings that have been documented in the literature.

3. Methods for Feature Extraction

Examine the literature on feature extraction methods that are frequently combined with neural networks to recognize speech, such as Mel Frequency Cepstral Coefficients (MFCC).LPC stands for linear predictive coding.Spectral characteristics.Compare the performance of different feature extraction algorithms in terms of their capacity to capture important information from voice signals.

4. Techniques for Training and Optimization

Provide an overview of earlier research that investigated different approaches to training and optimization for neural network-based speech recognition systems, such as gradient descent-based optimization algorithms and backpropagation.

strategies for regularization to avoid overfitting.Model selection and adjustment of hyperparameters.Emphasize new methods or developments in algorithm training that have been documented in the literature recently.

FLOW DIAGRAM:

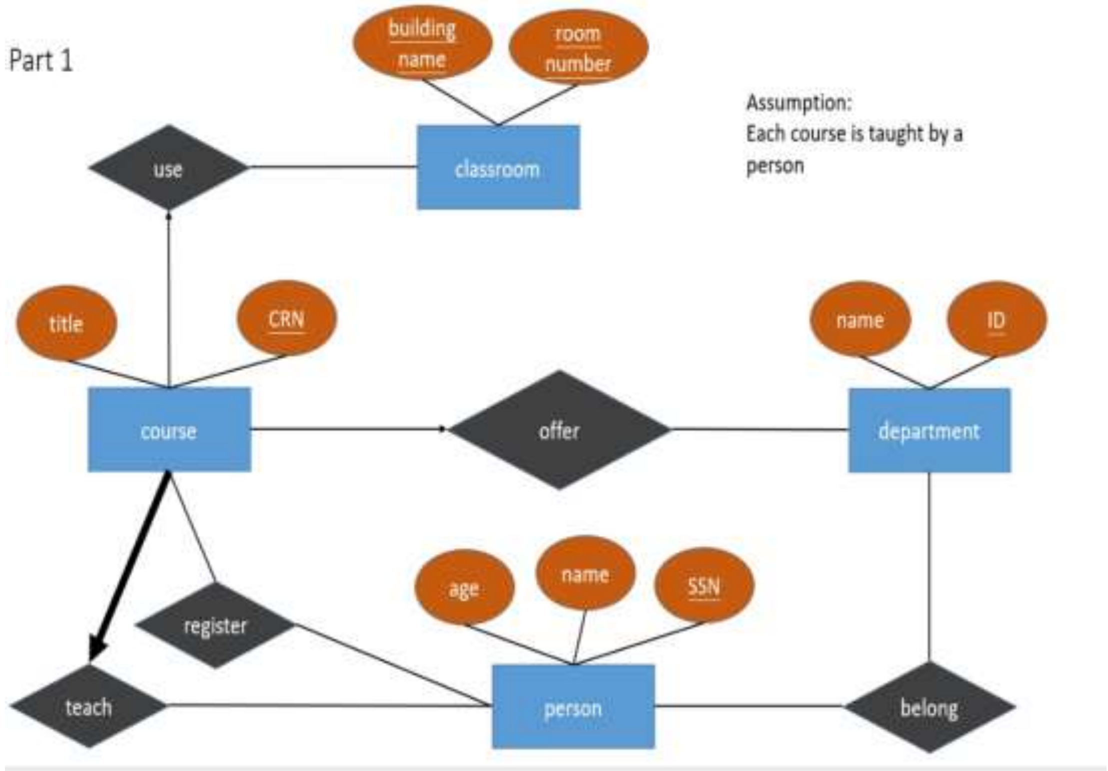

Fig 1. Flow Diagram

METHODOLOGY

The methodology for utilizing Artificial Neural Networks (ANNs) in speech recognition typically involves several key steps:

1. Data Collection and Preprocessing

- Acquire a large dataset of speech samples covering various speakers, accents, and environmental conditions.

- Preprocess the speech data to enhance quality, remove noise, and normalize features. This may involve techniques like noise reduction, endpoint detection, and feature extraction.

2. Feature Extraction

- Extract relevant features from the preprocessed speech signals. Common techniques include Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), and spectral features.

- Ensure that the extracted features capture the essential characteristics of speech while reducing dimensionality and computational complexity.

3. Neural Network Architecture Selection

- Choose an appropriate neural network architecture for the speech recognition task. Common architectures include:

  - Convolutional Neural Networks (CNNs) for feature extraction from spectrograms or other time-frequency representations.

  - Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU), for modeling temporal dependencies in speech sequences.

  - Hybrid architectures combining CNNs and RNNs for both feature extraction and sequence modeling.

4. Training

- Split the preprocessed data into training, validation, and test sets.

- Train the selected neural network architecture using the training data. This involves optimizing the network parameters (weights and biases) using backpropagation and gradient descent-based optimization algorithms.

- Monitor the performance on the validation set to prevent overfitting and fine-tune hyperparameters accordingly.

5. Evaluation

- Evaluate the trained model's performance on the test set using appropriate metrics such as accuracy, precision, recall, and F1-score.

- Analyze any misclassifications or errors to identify areas for improvement.

6. Deployment and Integration

- Deploy the trained neural network model into production environments for real-time or batch processing of speech inputs.

- Integrate the speech recognition system with other components of the application or system, such as natural language understanding modules or user interfaces.

7. Continuous Improvement

- Continuously collect new data and update the model to adapt to changes in speech patterns, environments, or user requirements.

- Explore advanced techniques, architectures, or algorithms to further improve the performance and robustness of the speech recognition system.

By following this methodology, researchers and practitioners can effectively harness the power of Artificial Neural Networks for speech recognition tasks, enabling more accurate and reliable human-computer interaction.

4. RESULTS:

1. Experimental Configuration

Give a summary of the experimental setting, mentioning specifics like the dataset that was used, the preprocessing methods that were used, and the neural network designs that were put into place.

Please provide any setups or hyperparameters unique to your investigations.

2. Metrics for Performance Evaluation:

Specify the evaluation criteria that will be used to gauge how well your neural network models perform in speech recognition applications.

Word error rate (WER), phoneme error rate (PER), recall, accuracy, precision, and F1-score are examples of common measurements.

3. Comparing the baseline:

Evaluate your neural network models' performance in comparison to baseline or cutting-edge techniques currently in use.

Emphasize any enhancements or benefits that your suggested approaches have brought about.

4. Measurable Outcomes:

Provide your tests' quantitative results in a tabular or graphical manner.

Incorporate performance indicators such as confusion matrices, loss curves, accuracy, and other pertinent data.

Give comparisons of the various neural network designs, feature extraction methods, or optimization approaches you used in your research.

5. Analysis of Qualitative Data:

Analyze your findings qualitatively, concentrating on particular instances or case studies to highlight the strengths and weaknesses of your neural network models.

Talk about any outliers, trends, or patterns in the data that you have noticed that could help you draw conclusions.

4. CONCLUSION:

1. Synopsis of Results:

Give a succinct overview of the key conclusions and outcomes that your work presents.

Give a brief overview of the research's experimental design, neural network designs, feature extraction methods, and performance assessment criteria.

2. Contributions:

Emphasize how your work has advanced the field of artificial neural network-based speech recognition.

Highlight any original methods, strategies, or discoveries that resulted from your research.

3. Significance & Implications:

Talk about how your findings might affect related study fields and the larger field of voice recognition.

Think about how your study adds to our understanding of the world today, tackles current issues, or creates new avenues for research.

4. Restrictions and Difficulties:

Recognize any restrictions or limits that you ran against when conducting your research, such as a lack of data, limited computational resources, or experimental assembly.

Think back on the difficulties encountered and the knowledge gained from conquering them.

5. Prospective Routes:

Provide possible directions for further investigation in light of your study's findings.

Determine open issues, uncharted research territory, or chances for additional advancements in artificial neural network-based voice recognition.

Talk about the potential effects of technological, methodological, or dataset developments on the field's future research paths.

Reference:

[1] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Network used for Speech Recognition", Journal of Automatic Control, University of Belgrade, Vol. 20, pp. 1-7, 2010.http://dx.doi.org/10.2298/JAC1001001G

[2] Vimal Krishnan VR, Athulya Jayakumar, Babu Anto P, "Speech Recognition of Isolated Malyalam Words Using Wavelet Feature and Artificial Neural Networks", 4th IEEE International Symposium on Electronic Design, Test and Application, 2008. http://dx.doi.org/10.1109/DELTA.2008.88

[3] Ganesh Tiwari,"Text Prompted Remote Speaker Authentication: Joint Speech & Speaker Recognition/Verification System.

[4] http://www.guidogybels.eu/asrp4.html

[5] Yashwanth H, Harish Mahendrakar and Suman Davia, " Automatic Speech recognition Using Audio Visual Cues", IEEE India Annual Conferencec pp. 166-169, 2004.

[6] G. Saha, Sandipan Chakroborty, Suman Senapati, "A New Silence Removal and Endpoint Deletion Algorithm for Speech and Speaker Recognition Applications.

[7] Urmila Shrawankar, Dr. Vilas Thakare, "Techniques for Feature Extraction in Speech Recognition System: A Comparative Study.

[8] Lei Xie, Zhi-Qiang Liu, "A Comparative Study of Audio Feature for Audio Visual Conversion in MPEG-4 Compliant Facial Animation",

Proc. of ICMLC Dalian, 13-16, August 2006.http://dx.doi.org/10.1109/icmlc.2006.259085

[9] Honig, Florian Stemmer, George Hacker, Christian Brugnara, Fabio, "Revising Perceptual Linear Prediction", In interspeech – 2005, pp. 2997-3000